

Joint Channel-and-Data Estimation for Large-MIMO Systems with Low-Precision ADCs

Chao-Kai Wen, Shi Jin, Kai-Kit Wong, Chang-Jen Wang, and Gang Wu

Abstract—The use of low precision (e.g., 1 – 3 bits) analog-to-digital converters (ADCs) in very large multiple-input multiple-output (MIMO) systems is a technique to reduce cost and power consumption. In this context, nevertheless, it has been shown that the training duration is required to be *very large* just to obtain an acceptable channel state information (CSI) at the receiver. A possible solution to the quantized MIMO systems is joint channel-and-data (JCD) estimation. This paper first develops an analytical framework for studying the quantized MIMO system using JCD estimation. In particular, we use the Bayes-optimal inference for the JCD estimation and realize this estimator utilizing a recent technique based on approximate message passing. Large-system analysis based on the replica method is then adopted to derive the asymptotic performances of the JCD estimator. Results from simulations confirm our theoretical findings and reveal that the JCD estimator can provide a significant gain over conventional pilot-only schemes in the quantized MIMO system.

I. INTRODUCTION

Very large multiple-input multiple-output (MIMO) or “massive MIMO” systems [1] are widely considered as a key technology for 5G wireless communications networks [2, 3]. Such systems promote the use of a very large number of antennas at the base station (BS) (e.g., hundreds or thousands) to serve a number of user terminals (e.g., tens or hundreds) in the same time-frequency resource. Nonetheless, the high dimensionality greatly increases hardware cost and power consumption. This motivates the study of MIMO systems with very low precision (e.g., 1 – 3 bits) analog-to-digital converters (ADCs).

Several aspects of low precision ADCs have been studied in the literature for single-input single-output (SISO) channels [4] and more recently MIMO channels [5] and references therein. In this paper, our focus is on signal detection for the quantized MIMO systems where each receiving antenna is equipped with a very low precision ADC. Prior work in this direction covered code-division multiple-access (CDMA) systems [6], massive MIMO systems [5–9], distributed antenna systems (DASs) [10], and compressed sensing [11]. However, most previous work assumed perfect channel state information at the receiver (CSIR). In particular, [8] revealed that in a MIMO system with one-bit ADC, to achieve the same performance as the full CSI case we have to use a very long training sequence (above 50

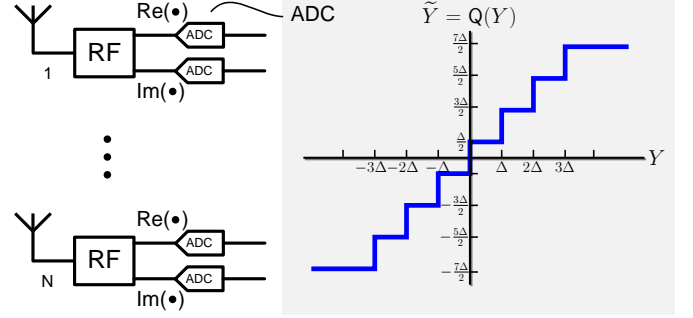


Fig. 1. The quantized MIMO system.

times the number of users). Clearly, the assumption of perfect CSIR becomes quite controversial particularly for quantized MIMO systems. The requirement of long training sequence motivates us to consider joint channel-and-data (JCD) estimation in which estimated payload data are utilized to aid channel estimation. A major advantage of JCD estimation is that relatively few pilot symbols are required to achieve the equalization channel and data estimation performances.

Although performance enhancement by using this technique is expected, we are not aware of any study for the *quantized* MIMO systems using JCD estimation.¹ In this paper, we take the important first step to analyze the achievable performance of the quantized MIMO system using JCD estimation. To this end, we use the Bayes-optimal inference for JCD estimation as this approach provides the minimal mean-square-error (MSE) with respect to (w.r.t.) the channels and payload data. However, the complexity for carrying out the Bayes-optimal JCD estimator appears prohibitive. To address this issue, we use a variant version of belief propagation (BP) to approximate the marginal distributions of each data and channel components. In particular, we modify the bilinear generalized approximate message passing (BiG-AMP) scheme in [15] and adapt it to the quantized MIMO system. We refer to the proposed method as GAMP-based JCD. Furthermore, by applying large-system analysis based on the replica method from statistical physics, we provide the theoretical performances for the Bayes-optimal JCD estimator.² Simulations are used to verify the efficiency of the proposed algorithm and the accuracy of our analysis.

¹In the context of *unquantized* MIMO system, several aspects of the JCD estimation have already been widely studied, see e.g., [12–14].

²In this paper, the Bayes-optimal JCD estimator is regarded as the *theoretical* optimal estimator, while the GAMP-based JCD algorithm can be thought of as a *practical* algorithm to approximate the theoretical optimal estimator.

C.-K. Wen and C.-J. Wang are with the Institute of Communications Engineering, National Sun Yat-sen University, Kaohsiung, Taiwan (e-mail: chaokai.wen@mail.nsysu.edu.tw). S. Jin is with the National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, P. R. China. K. Wong is with the Department of Electronic and Electrical Engineering, University College London, UK. G. Wu is with the National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, Chengdu 611731, P. R. China.

II. SYSTEM MODEL

We consider a block-fading uplink channel with K transmit antennas and N receive antennas, in which the channel remains constant over T consecutive symbol-intervals (i.e., a block). The received signal $\mathbf{Y} \in \mathbb{C}^{N \times T}$ over the block interval can be written in matrix form as

$$\mathbf{Y} = \frac{1}{\sqrt{K}} \mathbf{H} \mathbf{X} + \mathbf{W} = \mathbf{Z} + \mathbf{W}, \quad (1)$$

where $\mathbf{X} \in \mathbb{C}^{K \times T}$ denotes the transmit symbols in the block, $\mathbf{H} \in \mathbb{C}^{N \times K}$ is the matrix containing the fading coefficients associated to the channels between the transmit antennas and the receive antennas, $\mathbf{W} \in \mathbb{C}^{N \times T}$ represents the additive temporally and spatially white Gaussian noise with zero mean and element-wise variance σ_w^2 , and we define $\mathbf{Z} \triangleq \frac{1}{\sqrt{K}} \mathbf{H} \mathbf{X}$.

In the quantized MIMO system, as shown in Figure 1, each received signal component Y^{ij} , $1 \leq i \leq N$, $1 \leq j \leq T$ are quantized separately into a finite set of prescribed values by a B-bit quantizer Q_c . The resulting quantized signals can read

$$\tilde{\mathbf{Y}} = Q_c(\mathbf{Z} + \mathbf{W}). \quad (2)$$

Specifically, each complex-valued quantizer $Q_c(\cdot)$ is defined as $\tilde{Y}^{ij} = Q_c(Y^{ij}) \triangleq Q(\text{Re}\{Y^{ij}\}) + jQ(\text{Im}\{Y^{ij}\})$, i.e., the real and imaginary parts are quantized separately. The real-valued quantizer Q maps a real-valued input to one of the 2^B bins, which are characterized by the set of $2^B - 1$ thresholds $[r_1, r_2, \dots, r_{2^B-1}]$, such that $-\infty < r_1 < r_2 < \dots < r_{2^B-1} < \infty$. For notational consistence, we define $r_0 = -\infty$ and $r_{2^B} = \infty$. The output \tilde{Y} is assigned a value in $(r_{b-1}, r_b]$ when the quantizer input Y falls in the interval $(r_{b-1}, r_b]$ (namely, the b -th bin). For example, the threshold of a typical uniform quantizer with the quantization step-size Δ is given by

$$r_b = (-2^{B-1} + b)\Delta, \quad \text{for } b = 1, \dots, 2^B - 1, \quad (3)$$

and the quantization output is assigned the value $r_b - \frac{\Delta}{2}$ when the input falls in the b -th bin. Figure 1 shows an example of the 3-bit uniform quantizer.

Since the channel matrix \mathbf{H} needs to be estimated at the receiver, we make the first T_1 symbols of the block of T symbols serve as pilot sequences. The remaining $T_2 = T - T_1$ symbols are used for data transmissions. This setting is equivalent to partitioning \mathbf{X} as $\mathbf{X} = [\mathbf{X}_1 \mathbf{X}_2]$ with $\mathbf{X}_1 \in \mathbb{C}^{K \times T_1}$ and $\mathbf{X}_2 \in \mathbb{C}^{K \times T_2}$. The training and data phases are referred to as t-phase and d-phase, respectively. We assume that the matrix \mathbf{X}_1 (or \mathbf{X}_2) is composed of independent and identically distributed (i.i.d.) random variables X_1 (X_2) generated from a known probability distribution P_{X_1} (or P_{X_2}), i.e.,

$$P_{\mathbf{X}}(\mathbf{X}) = P_{X_1}(\mathbf{X}_1)P_{X_2}(\mathbf{X}_2) \quad (4)$$

with $P_{X_1}(\mathbf{X}_1) = \prod_{i,j} P_{X_1}(X_1^{ij})$, $P_{X_2}(\mathbf{X}_2) = \prod_{i,j} P_{X_2}(X_2^{ij})$.

Since the pilot and data symbols should appear on constellation points uniformly, the ensemble averages of $\{X_1^{ij}\}$ and $\{X_2^{ij}\}$ are assumed to be zero. In addition, we let $\sigma_{x_1}^2$ and $\sigma_{x_2}^2$ be the transmit powers during the t-phase and d-phase, respectively, i.e., $E\{|X_1^{ij}|^2\} = \sigma_{x_1}^2$ and $E\{|X_2^{ij}|^2\} = \sigma_{x_2}^2$.

Similarly, we assume that each entry H^{ij} is drawn from the complex Gaussian distribution $\mathcal{N}_{\mathbb{C}}(0, \sigma_h^2)$, where σ_h^2 indicates the large scale fading factor.³ Let $P_{\mathbf{H}}(H^{ij}) \equiv \mathcal{N}_{\mathbb{C}}(0, \sigma_h^2)$. Then we have

$$P_{\mathbf{H}}(\mathbf{H}) = \prod_{ij} P_{\mathbf{H}}(H^{ij}). \quad (5)$$

III. JCD ESTIMATION

Our focus is on the setting where the receiver knows the distributions of \mathbf{H} and \mathbf{X} but not their realizations. In the conventional pilot-only scheme, the receiver first uses \mathbf{Y}_1 and \mathbf{X}_1 to generate an estimate of \mathbf{H} and then uses the estimated channel for estimating the data \mathbf{X}_2 from \mathbf{Y}_2 [8]. In contrast to the pilot-only scheme, we consider JCD estimation, where the BS estimates both \mathbf{H} and \mathbf{X}_2 from $\tilde{\mathbf{Y}}$ given \mathbf{X}_1 . We treat the problem in the framework of Bayesian inference [16].

To this end, we first define the likelihood, which is the distribution of the received signals under (2) conditional on the unknown parameters, as:

$$P(\tilde{\mathbf{Y}}|\mathbf{H}, \mathbf{X}) = \prod_{i=1}^N \prod_{j=1}^T P_{\text{out}}(\tilde{Y}^{ij} | Z^{ij}), \quad (6)$$

where

$$P_{\text{out}}(\tilde{Y} | Z) = \left(\frac{1}{\sqrt{\pi\sigma_w^2}} \int_{r_{b-1}}^{r_b} dy e^{-\frac{(y - \text{Re}(Z))^2}{\sigma_w^2}} \right) \times \left(\frac{1}{\sqrt{\pi\sigma_w^2}} \int_{r_{b'-1}}^{r_{b'}} dy e^{-\frac{(y - \text{Im}(Z))^2}{\sigma_w^2}} \right) \quad (7)$$

as $\text{Re}(\tilde{Y})$ and $\text{Im}(\tilde{Y})$ fall in the b -th and the b' -th bins, respectively, i.e., $\text{Re}(\tilde{Y}) = r_b$ and $\text{Im}(\tilde{Y}) = r_{b'}$. Let $\Phi(x) \triangleq \int_{-\infty}^x Dz$ with $Dz \triangleq \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$. Then (7) has the following closed-form expression

$$P_{\text{out}}(\tilde{Y} | Z) = \Psi_b(\text{Re}(Z)) \Psi_{b'}(\text{Im}(Z)), \quad (8)$$

where

$$\Psi_b(x) \triangleq \Phi\left(\frac{\sqrt{2}(r_b - x)}{\sigma_w}\right) - \Phi\left(\frac{\sqrt{2}(r_{b-1} - x)}{\sigma_w}\right). \quad (9)$$

The prior distributions of \mathbf{H} and \mathbf{X} are given by (5) and (4), respectively. Then the posterior probability can be computed according to Bayes' rule:

$$P(\mathbf{H}, \mathbf{X} | \tilde{\mathbf{Y}}) = \frac{P(\tilde{\mathbf{Y}}|\mathbf{H}, \mathbf{X}) P_{\mathbf{H}}(\mathbf{H}) P_{\mathbf{X}}(\mathbf{X})}{P(\tilde{\mathbf{Y}})}, \quad (10)$$

where $P(\tilde{\mathbf{Y}}) = \int_{\mathbf{H}} \int_{\mathbf{X}} d\mathbf{H} d\mathbf{X} P(\tilde{\mathbf{Y}}|\mathbf{H}, \mathbf{X}) P_{\mathbf{H}}(\mathbf{H}) P_{\mathbf{X}}(\mathbf{X})$ is the marginal likelihood.

Given the posterior probability, an estimate for X^{ij} can be obtained by the posterior mean

$$\hat{X}^{ij} = \int dX^{ij} \mathcal{P}(X^{ij}) X^{ij}, \quad (11)$$

³For ease of notation, we consider the case where all the transmits have the same large-scale fading factor but our results can be easily extended.

where

$$\mathcal{P}(X^{ij}) = \int_{\mathbf{H}} \int_{\mathbf{X} \setminus X^{ij}} d\mathbf{H} d\mathbf{X} P(\mathbf{H}, \mathbf{X} | \tilde{\mathbf{Y}}) \quad (12)$$

is the marginal posterior probability of X^{ij} . Here, the notation $\int_{\mathbf{X} \setminus X^{ij}} d\mathbf{X}$ denotes the integration over all the variables in \mathbf{X} except for X^{ij} . If the MSE of an estimate $\hat{\mathbf{X}}$ w.r.t. \mathbf{X} is defined as

$$\text{mse}(\mathbf{X}_t | \tilde{\mathbf{Y}}) = \int_{\mathbf{H}} \int_{\mathbf{X}} d\mathbf{H} d\mathbf{X} P(\mathbf{H}, \mathbf{X} | \tilde{\mathbf{Y}}) \|\hat{\mathbf{X}}_t - \mathbf{X}_t\|_F^2, \quad (13)$$

for $t = 1, 2$, then the posterior mean estimator (11) gives the minimum MSE (MMSE) [16]. Notice that given a *known* pilot matrix \mathbf{X}_1 , i.e., $P_{\mathbf{X}_1}(\mathbf{X}_1) = \delta(\mathbf{X}_1 - \mathbf{X}_1)$, we can easily obtain $\hat{\mathbf{X}}_1^{ij} = \mathbf{X}_1^{ij}$ from (11). In this case, we have $\text{mse}(\mathbf{X}_1 | \tilde{\mathbf{Y}}) = 0$.

Similarly, the Bayes estimate of H^{ij} is given by

$$\hat{H}^{ij} = \int dH^{ij} \mathcal{P}(H^{ij}) H^{ij} \quad (14)$$

where $\mathcal{P}(H^{ij}) = \int_{\mathbf{H} \setminus H^{ij}} \int_{\mathbf{X}} d\mathbf{H} d\mathbf{X} P(\mathbf{H}, \mathbf{X} | \tilde{\mathbf{Y}})$ denotes the marginal posterior probability of H^{ij} . The estimate \hat{H} also minimizes the MSE

$$\text{mse}(\mathbf{H} | \tilde{\mathbf{Y}}) = \int_{\mathbf{H}} \int_{\mathbf{X}} d\mathbf{H} d\mathbf{X} P(\mathbf{H}, \mathbf{X} | \tilde{\mathbf{Y}}) \|\hat{\mathbf{H}} - \mathbf{H}\|_F^2. \quad (15)$$

Hereafter, we will refer to (11) and (14) as the Bayes-optimal estimator.

Although the Bayes-optimal estimator provides the MMSE estimates, direct computations of (11) and (14) are intractable due to high-denominational integrals involved in the marginal posteriors $\mathcal{P}(X^{ij})$ and $\mathcal{P}(H^{ij})$. In [17], BP provides a practical alternative to approximate these marginal posteriors. In the recent compressed sensing literature, the Bayesian framework in combination with a BP algorithm has given rise to the so-called approximate message passing (AMP) algorithm [18] and the generalized AMP (GAMP) [19]. Applying this development to our context of the MIMO system means that when \mathbf{H} is perfectly known, GAMP can provide a tractable way to approximate the marginal posteriors $\mathcal{P}(X^{ij})$'s. Remarkably, it has proved that the approximations become exact in the large-system limit for dense matrix \mathbf{H} with sub-Gaussian entries. More recently, Parker *et al.* in [15] applied the same strategy of GAMP to the problem of reconstructing matrices from bilinear noisy observations (i.e., reconstructing \mathbf{H} and \mathbf{X} from \mathbf{Y}), which is referred to as bilinear GAMP (BiG-AMP). The BiG-AMP scheme can be applied to tackle the Bayes-optimal JCD estimator and we can adapt it to be used in the quantized MIMO setting. We call the developed algorithm GAMP-based JCD algorithm. Due to space limitations, we remove details of the algorithm development in this paper while we will show its simulation performances later in Section V.

IV. PERFORMANCE ANALYSIS

knowing the theoretical lower bound of the estimate is useful to assess any developed algorithm. Therefore, in this section, our objective is to derive analytical results for the average MSEs of \mathbf{X}_2 and \mathbf{H} for the Bayes-optimal JCD estimator,

i.e., $\text{mse}_{X_t} \triangleq E\{\text{mse}(\mathbf{X}_t | \tilde{\mathbf{Y}})\}$ and $\text{mse}_H \triangleq E\{\text{mse}(\mathbf{H} | \tilde{\mathbf{Y}})\}$. Our analysis investigates the high-dimensional regime where $N, K, T \rightarrow \infty$ but the ratios $N/K = \alpha$, $T/K = \beta$, $T_t/K = \beta_t$, for $t = 1, 2$ are fixed and finite. For convenience, we simply use $K \rightarrow \infty$ (or refer to as the large-system limit) to denote this high-dimension limit. Following the argument of [20, 21], it can be shown that mse_{X_t} and mse_H are saddle points of the average free entropy

$$\Phi \triangleq \frac{1}{K^2} E_{\tilde{\mathbf{Y}}} \left\{ \log P(\tilde{\mathbf{Y}}) \right\}, \quad (16)$$

where $P(\tilde{\mathbf{Y}})$ denotes the marginal likelihood in (10), namely the partition function. The major difficulty in computing (16) is the expectation of the logarithm of $P(\tilde{\mathbf{Y}})$, which, nevertheless, can be facilitated by rewriting Φ as [22]

$$\Phi = \frac{1}{K^2} \lim_{\tau \rightarrow 0} \frac{\partial}{\partial \tau} \log E_{\tilde{\mathbf{Y}}} \left\{ P^\tau(\tilde{\mathbf{Y}}) \right\}. \quad (17)$$

The expectation operator is moved inside the log-function. We first evaluate $E_{\tilde{\mathbf{Y}}} \{P^\tau(\tilde{\mathbf{Y}})\}$ for an integer-valued τ , and then generalize the result to any positive real number τ . This technique is called the replica method, and has been widely adopted in the field of statistical physics [22] and information theory literature, e.g., [12, 23–29]. Under the assumption of replica symmetry (RS), the following results are obtained.

Proposition 1: As $K \rightarrow \infty$, the asymptotic MSEs w.r.t. \mathbf{X}_t and \mathbf{H} are associated with the MSEs for the scalar Gaussian channels:

$$Y_{\tilde{q}_H} = \sqrt{\tilde{q}_H} H + W_H, \quad (18)$$

$$Y_{\tilde{q}_{X_t}} = \sqrt{\tilde{q}_{X_t}} X_t + W_{X_t}, \quad (19)$$

where $W_H, W_{X_t} \sim \mathcal{N}_{\mathbb{C}}(0, 1)$ are independent of $H \sim P_H$ and $X_t \sim P_{X_t}$. Here, the parameters \tilde{q}_H and \tilde{q}_{X_t} are the solutions to the set of fixed-point equations

$$\tilde{q}_H = \sum_{t=1}^2 \beta_t q_{X_t} \chi_t, \quad (20a)$$

$$\tilde{q}_{X_t} = \alpha q_H \chi_t, \quad (20b)$$

$$q_H = c_H - \text{mse}_H, \quad (20c)$$

$$q_{X_t} = c_{X_t} - \text{mse}_{X_t}. \quad (20d)$$

where $\text{mse}_H = E\{|H - E\{H | Y_{\tilde{q}_H}\}|^2\}$ and $\text{mse}_{X_t} = E\{|X_t - E\{X_t | Y_{\tilde{q}_{X_t}}\}|^2\}$ are the asymptotic MSEs w.r.t. \mathbf{X}_t and \mathbf{H} , respectively, $c_{X_t} \triangleq E\{|X_t|^2\} = \sigma_{x_t}^2$, $c_H \triangleq E\{|H|^2\} = \sigma_h^2$, and

$$\chi_t \triangleq \sum_{b=1}^{2^B-1} \int Dv \frac{\left(\Psi'_b(\sqrt{q_H q_{X_t}} v) \right)^2}{\Psi_b(\sqrt{q_H q_{X_t}} v)}, \quad (21)$$

with

$$\Psi_b(V_t) \triangleq \Phi \left(\frac{\sqrt{2} r_b - V_t}{\sqrt{\sigma_w^2 + c_H c_{X_t} - q_H q_{X_t}}} \right) - \Phi \left(\frac{\sqrt{2} r_{b-1} - V_t}{\sqrt{\sigma_w^2 + c_H c_{X_t} - q_H q_{X_t}}} \right), \quad (22)$$

and

$$\begin{aligned}\Psi'_b(V_t) &\triangleq \frac{\partial \Psi_b(V_t)}{\partial V_t} \\ &= \frac{e^{-\frac{(\sqrt{2}r_b - V_t)^2}{2(\sigma_w^2 + c_H c_{X_t} - q_H q_{X_t})}} - e^{-\frac{(\sqrt{2}r_b - 1 - V_t)^2}{2(\sigma_w^2 + c_H c_{X_t} - q_H q_{X_t})}}}{\sqrt{2\pi(\sigma_w^2 + c_H c_{X_t} - q_H q_{X_t})}}.\end{aligned}\quad (23)$$

In the t-phase, i.e., $t = 1$, the pilot matrix \mathbf{X}_1 is known. Thus, we substitute $\text{mse}_{X_1} = 0$ into the above expressions.

Proof: An outline proof is given in the appendix. ■

The above result reveals that in the large-system limit, the performance of the quantized MIMO system employing the Bayes-optimal JCD estimator can be fully characterized by the equivalent scalar Gaussian channels (18) and (19). For example, the achievable rate under the separate decoding (SD) is the mutual information between $Y_{\tilde{q}_{X_t}}$ and X_t for the scalar Gaussian channel (19). Note that in contrast to joint detection and decoding, the SD involves the joint multiuser detection followed by a bank of independent decoders. Also, the corresponding MSE w.r.t. \mathbf{H} can be evaluated through the scalar Gaussian channel (18). Specifically, if the signal is drawn from a quadrature phase shift keying (QPSK) constellation, the corresponding bit error rate (BER) reads

$$P_e = \mathcal{Q}(\sqrt{\tilde{q}_X}), \quad (24)$$

where $\mathcal{Q}(x) \triangleq \int_x^\infty \mathcal{D}z$ is the Q function, the MSE w.r.t. payload data \mathbf{X}_2 is given by

$$\text{mse}_{X_2} = 1 - \int \mathcal{D}z \tanh(\tilde{q}_{X_2} + \sqrt{\tilde{q}_{X_2}}z), \quad (25)$$

and the corresponding MSE w.r.t. \mathbf{H} is $\text{mse}_H = \frac{\sigma_h^2}{1 + \sigma_h^2 \tilde{q}_H}$.

If the channel matrix \mathbf{H} is perfectly known, the t-phase is not required so $\beta_2 = \beta$ and $\beta_1 = 0$. Since there is only one phase in \mathbf{X} , we omit the phase indices (t) from all the concerned parameters in this case. Because \mathbf{H} is perfectly known, we set $\text{mse}_H = 0$. Plugging this into (20c), we immediately obtain $q_H = c_H = \sigma_h^2$, which leads to $q_H q_X = c_H q_X$, and $c_H c_X - q_H q_X = c_H(c_X - q_X) = c_H \text{mse}_X$. It turns out that the equivalent signal-to-interference-plus-noise ratio (SINR) of the scalar Gaussian channel (19) is given by

$$\tilde{q}_X = \alpha c_H \left(\sum_{b=1}^{2^B-1} \int \mathcal{D}v \frac{(\Psi'_b(\sqrt{c_H q_X v}))^2}{\Psi_b(\sqrt{c_H q_X v})} \right), \quad (26)$$

and the asymptotic MSE w.r.t. \mathbf{X} is given by $\text{mse}_X = \mathbb{E}\{|X - \mathbb{E}\{X|Y_{\tilde{q}_X}\}|^2\}$. If QPSK is used, the MSE in (25) together with \tilde{q}_X in (26) agree with [6, (7) & (8)]. More precisely, in [6], the real-valued system with BPSK signal is considered. In this case, $\sqrt{2}r_b$ in our paper should be replaced by r_b .

V. NUMERICAL RESULTS

To verify the accuracy of our analytical results, we compare the analytical BER expression (24) with that obtained by simulations (performed by the GAMP-based JCD algorithm) under the quantized MIMO system with QPSK constellation.

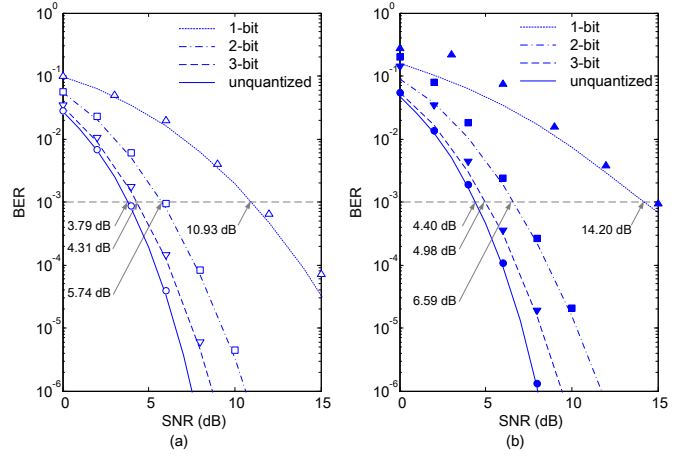


Fig. 2. BER versus $\text{SNR} = 1/\sigma_w^2$ for QPSK constellations. In the results, the JCD estimation scheme is used under the settings with a) perfect CSIR and b) no CSIR. Curves denote analytical results and markers denote Monte-Carlo simulation results achieved by the GAMP-based JCD algorithm.

The simulation results are obtained by averaging over 10,000 channel realizations. The parameters of the system are set as follows: $K = 50$, $N = 200$, $T_1 = 50$, and $T_2 = 450$. The pilot sequences of length T_1 are randomly generated. In all the following simulations, we use the typical uniform quantizer with the quantization step-size $\Delta = \sqrt{0.25}$. Note that we do not optimize the quantization step-size but select a good one for general scenarios. We leave the related issue to our future work. Figure 2 shows the corresponding BERs results for the cases of 1) perfect CSIR and b) no CSIR. We observe that the analytical BER expression (24) generally predicts well the behavior of the GAMP-based JCD algorithm. For the case with no CSIR, the GAMP-based JCD algorithm cannot work as well as that predicted by the analytical result at low SNRs. This would be because the GAMP-based JCD algorithm is only an approximation to the Bayes-optimal JCD estimator. This gap has motivated the search for other improved estimators in the future. In addition, from Figure 2, we see that the performance degradation due to 3-bit quantization is small. For instance, if we target the SNR to that attained by the unquantized system with perfect CSIR at $\text{BER} = 10^{-3}$, the 3-bit Bayes-optimal JCD estimator only incurs a loss of 1.19 dB. Even with 2-bit quantization, the loss of 2.8 dB remains acceptable.

Comparing Figures 2(a) and 2(b), we see that the loss due to no CSIR is small for the proposed JCD estimator. Therefore, it is of interest to evaluate the improvement due to the JCD estimation. Following the same system parameters as before, Figure 3 compares the BERs under the conventional pilot-only scheme and the proposed JCD estimation scheme. For the pilot-only scheme, we adopt the receiver structure of [8], which employs the least squares (LS) channel estimate for the quantized MIMO system. However, unlike [8], we then employ the GAMP algorithm for payload data detection based on the estimated channel. Therefore, the BERs of the pilot-only scheme shown in Figure 3 are expected to be better than

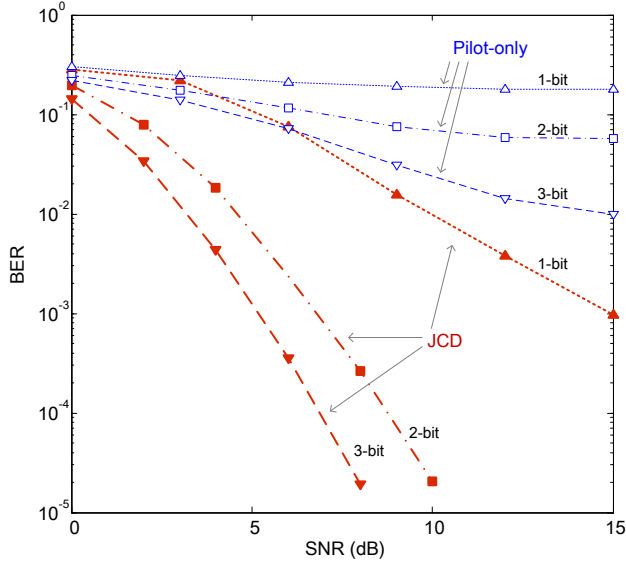


Fig. 3. Average BER versus SNR for QPSK constellations. In the results, the GAMP-based JCD algorithm and pilot-only scheme are used. Plots are based on Monte-Carlo simulation results.

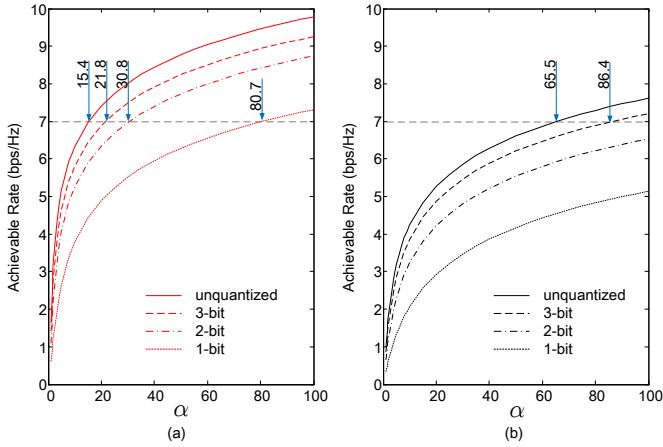


Fig. 4. The achievable rates as functions of $\alpha = N/K$ for a) the Bayes-optimal JCD estimator and b) the pilot-only scheme. $\beta = 10$, $\beta_1 = 1$, $\sigma_w^2 = 10^{-1}$, and $X_2^{ij} \sim \mathcal{N}_{\mathbb{C}}(0, 1)$.

that employing suboptimal criteria [8]. Even so, as can be seen from Figure 3, the JCD estimation still shows a large improvement over the pilot-only scheme.

Finally, we compare the achievable rates as functions of the antenna ratio $\alpha = N/K$ for the Bayes-optimal JCD estimator and the pilot-only scheme under different quantization precisions in Figure 4. Note that unlike the QPSK signals used in pervious simulations, we consider Gaussian signal, i.e., $X_2 \sim \mathcal{N}_{\mathbb{C}}(0, 1)$, in this experiment. We observe that the achievable rates of all the quantization precisions increase as the receive antenna numbers even for the 1-bit receivers. This implies that the use of high-order modulation schemes is also possible in 1-bit MIMO systems, which shares the same view as [5]. This property is quite different from the quantized SISO

system, where the achievable rate is always upper bounded by B bits in a B-bit receiver [4]. If we fix the achievable rate to 7 bps/Hz, the number of receive antennas for the 3-bit Bayes-optimal JCD estimator is only about $\frac{21.8}{15.4} \approx 1.4$ times of the unquantized Bayes-optimal JCD estimator (the benchmark receiver), while even with unquantized receivers, the number of receive antennas for the pilot-only scheme requires about $\frac{65.5}{15.4} \approx 4.3$ times of the benchmark receiver. The penalty of increasing 1.4 times of antenna numbers with very low precision ADCs seems to be quite acceptable.

VI. CONCLUSION

We have developed a framework for studying the best possible estimation performance of the quantized MIMO system. In particular, we used the Bayes-optimal inference for the JCD estimation and realized this estimation by applying the BiG-AMP technique. Additionally, the asymptotic performances (e.g., MSEs) w.r.t. the channels and the payload data are derived in the large-system limit. A set of Monte-Carlo simulations was conducted to illustrate that our analytical results provide an accurate prediction for the performances of the Bayes-optimal JCD estimator. The numerical results have also revealed that the JCD estimation scheme provides tremendous improvement over the conventional pilot-only scheme.

APPENDIX A: PROOF OF PROPOSITION 1

As stated in Section IV, the MSEs of interest are saddle points of the average free entropy (16). However, direct calculation is very difficult. Thus, we resort to the replica method by computing the replicate partition function $E_{\tilde{\mathbf{Y}}} \{P^\tau(\mathbf{Y})\}$ in (17), which with the definition of (10) can be expressed as

$$E_{\tilde{\mathbf{Y}}} \{P^\tau(\mathbf{Y})\} = E_{\mathcal{H}, \mathcal{X}} \left\{ \int d\tilde{\mathbf{Y}} \prod_{a=0}^{\tau} P_{\text{out}}(\tilde{\mathbf{Y}} | \mathbf{Z}^{(a)}) \right\}, \quad (27)$$

where we define $\mathbf{Z}^{(a)} \triangleq \mathbf{H}^{(a)} \mathbf{X}^{(a)} / \sqrt{K}$ with $\mathbf{H}^{(a)}$ and $\mathbf{X}^{(a)}$ being the a -th replica of \mathbf{H} and \mathbf{X} respectively, and the notations $\mathcal{X} \triangleq \{\mathbf{X}^{(a)}, \forall a\}$ and $\mathcal{H} \triangleq \{\mathbf{H}^{(a)}, \forall a\}$. Here, $(\mathbf{H}^{(a)}, \mathbf{X}^{(a)})$ are random matrices taken from the distribution $(P_{\mathbf{H}}, P_{\mathbf{X}})$ for $a = 0, 1, \dots, \tau$. In addition, $\int d\tilde{\mathbf{Y}}$ denotes the integral w.r.t. a discrete measure because the quantized output $\tilde{\mathbf{Y}}$ is a finite set. Next, we will focus on calculating the right-hand side of (27), which can be done by applying the techniques in [14, 20, 21] after additional manipulations.

First, in order to average over $(\mathcal{H}, \mathcal{X})$, we introduce two $(\tau + 1) \times (\tau + 1)$ matrices $\mathbf{Q}_H = [Q_H^{ab}]$ and $\mathbf{Q}_{X_t} = [Q_{X_t}^{ab}]$ whose elements are defined by $Q_H^{ab} = \mathbf{h}_n^{(b)} (\mathbf{h}_n^{(a)})^\dagger / K$ and $Q_{X_t}^{ab} = (\mathbf{x}_j^{(a)})^\dagger \mathbf{x}_j^{(b)} / K$. Here, $\mathbf{h}_n^{(a)}$ denotes the n th row vector of $\mathbf{H}^{(a)}$, and $\mathbf{x}_j^{(a)}$ denotes the j th column vector of $\mathbf{X}^{(a)}$ corresponding to phase block t , and \mathcal{T}_t for $t = 1, 2$ represents the set of all symbol indices in phase block t . The definitions

of \mathbf{Q}_H and \mathbf{Q}_{X_t} are equivalent to

$$1 = \int \prod_{n=1}^N \prod_{0 \leq a \leq b}^{\tau} \delta \left(\mathbf{h}_n^{(b)} (\mathbf{h}_n^{(a)})^\dagger - K Q_H^{ab} \right) dQ_H^{ab},$$

$$1 = \int \prod_{t=1}^2 \prod_{j \in \mathcal{T}_t} \prod_{0 \leq a \leq b}^{\tau} \delta \left((\mathbf{x}_j^{(a)})^\dagger \mathbf{x}_j^{(b)} - K Q_{X_t}^{ab} \right) dQ_{X_t}^{ab},$$

where $\delta(\cdot)$ denotes Dirac's delta. Let $\mathbf{Q}_X \triangleq \{\mathbf{Q}_{X_t}, \forall t\}$ and $\mathbf{Z} \triangleq \{\mathbf{Z}^{(a)}, \forall a\}$. Inserting the above into (27) yields

$$\mathbb{E}_{\tilde{\mathbf{Y}}} \{P^\tau(\tilde{\mathbf{Y}})\} = \int e^{K^2 \mathcal{G}^{(\tau)}} d\mu_H^{(\tau)}(\mathbf{Q}_H) d\mu_X^{(\tau)}(\mathbf{Q}_X), \quad (28)$$

where

$$\mathcal{G}^{(\tau)}(\mathbf{Q}_Z) \triangleq \frac{1}{K^2} \log \mathbb{E}_{\mathbf{Z}} \left\{ \int d\tilde{\mathbf{Y}} \prod_{a=0}^{\tau} P_{\text{out}}(\tilde{\mathbf{Y}} | \mathbf{Z}^{(a)}) \right\},$$

$$\mu_H^{(\tau)}(\mathbf{Q}_H) \triangleq \mathbb{E}_{\mathbf{H}} \left\{ \prod_{n,a,b} \delta \left(\mathbf{h}_n^{(b)} (\mathbf{h}_n^{(a)})^\dagger - K Q_H^{ab} \right) \right\},$$

$$\mu_X^{(\tau)}(\mathbf{Q}_X) \triangleq \mathbb{E}_{\mathbf{X}} \left\{ \prod_{t,j,a,b} \delta \left((\mathbf{x}_j^{(a)})^\dagger \mathbf{x}_j^{(b)} - K Q_{X_t}^{ab} \right) \right\}.$$

Then using the Fourier representation of the δ function and computing the integrals by the saddle point method, we attain

$$\frac{1}{K^2} \mathbb{E}_{\tilde{\mathbf{Y}}} \{P^\tau(\tilde{\mathbf{Y}})\} = \underset{\mathbf{Q}_H, \mathbf{Q}_X, \tilde{\mathbf{Q}}_H, \tilde{\mathbf{Q}}_X}{\text{Extr}} \{ \Phi^{(\tau)} \} \quad (29)$$

with

$$\Phi^{(\tau)} \triangleq \frac{1}{K^2} \log \mathbb{E}_{\mathbf{Z}} \left\{ \prod_{n,t,j \in \mathcal{T}_t} \int dy_{n,j} \prod_a P_{\text{out}}(y_{n,j} | z_{n,j}^{(a)}) \right\}$$

$$+ \frac{1}{K^2} \log \mathbb{E}_{\mathbf{H}} \left\{ \prod_n e^{\text{tr}(\tilde{\mathbf{Q}}_H \mathbf{H}_n^\dagger \mathbf{H}_n)} \right\} - \alpha \text{tr}(\tilde{\mathbf{Q}}_H \mathbf{Q}_H)$$

$$+ \frac{1}{K^2} \log \mathbb{E}_{\mathbf{X}} \left\{ \prod_t e^{\text{tr}(\tilde{\mathbf{Q}}_{X_t} \mathbf{X}_t^\dagger \mathbf{X}_t)} \right\} - \sum_t \beta_t \text{tr}(\tilde{\mathbf{Q}}_{X_t} \mathbf{Q}_{X_t}),$$

where $\text{Extr}_x \{f(x)\}$ represents the extreme value of $f(x)$ w.r.t. x , $\mathbf{Q}_H = [\mathbf{Q}_H^{ab}] \in \mathbb{C}^{(\tau+1) \times (\tau+1)}$ and $\mathbf{Q}_X \triangleq \{\mathbf{Q}_{X_t} = [\mathbf{Q}_{X_t}^{ab}] \in \mathbb{C}^{(\tau+1) \times (\tau+1)}, \forall t\}$. According to (17), the average free entropy turns out to be $\Phi = \lim_{\tau \rightarrow 0} \frac{\partial}{\partial \tau} \text{Extr}_{\mathbf{Q}_H, \mathbf{Q}_X, \tilde{\mathbf{Q}}_H, \tilde{\mathbf{Q}}_X} \{ \Phi^{(\tau)} \}$.

The saddle points of $\Phi^{(\tau)}$ can be obtained by seeking the point of zero gradient w.r.t. $\{\mathbf{Q}_H, \mathbf{Q}_{X_t}, \tilde{\mathbf{Q}}_H, \tilde{\mathbf{Q}}_{X_t}\}$. However, in doing so, it is prohibitive to get explicit expressions about the saddle points. Therefore, we assume that the saddle points follow the RS form [22] as $\mathbf{Q}_H = (c_H - q_H)\mathbf{I} + q_H\mathbf{1}$, $\tilde{\mathbf{Q}}_H = (\tilde{c}_H - \tilde{q}_H)\mathbf{I} + \tilde{q}_H\mathbf{1}$, $\mathbf{Q}_{X_t} = (c_{X_t} - q_{X_t})\mathbf{I} + q_{X_t}\mathbf{1}$, $\tilde{\mathbf{Q}}_{X_t} = (\tilde{c}_{X_t} - \tilde{q}_{X_t})\mathbf{I} + \tilde{q}_{X_t}\mathbf{1}$, where \mathbf{I} denotes the identity matrix and $\mathbf{1}$ denotes the all-one matrix. In addition, the application of the central limit theorem suggests that the $\mathbf{z}_{n,j} \triangleq [z_{n,j}^{(0)} z_{n,j}^{(1)} \cdots z_{n,j}^{(\tau)}]^T$ are Gaussian random vectors with $(\tau+1) \times (\tau+1)$ covariance matrix \mathbf{Q}_{Z_t} . If $j \in \mathcal{T}_t$, then the

(a, b) th entry of \mathbf{Q}_{Z_t} is given by

$$(z_{n,j}^{(a)})^* z_{n,j}^{(b)} = Q_H^{ab} Q_{X_t}^{ab} \triangleq Q_{Z_t}^{ab}. \quad (30)$$

Therefore, we set $\mathbf{Q}_{Z_t} = (c_H c_{X_t} - q_H q_{X_t})\mathbf{I} + q_H q_{X_t}\mathbf{1}$, which is equivalent to introduce to the Gaussian random variable $\mathbf{z}_{n,j}$ for $j \in \mathcal{T}_t$ as

$$z_{n,j}^{(a)} = \sqrt{c_H c_{X_t} - q_H q_{X_t}} u^{(a)} + \sqrt{q_H q_{X_t}} v \quad (31)$$

for $a = 0, 1, \dots, \tau$, where $u^{(a)}$ and v are independent standard complex Gaussian random variables.

Substituting these RS expression into $\Phi^{(\tau)}$ leads to the RS expression of $\Phi^{(\tau)}$. With the RS, we only have to determine the parameters $\{c_H, q_H, c_{X_t}, q_{X_t}, \tilde{c}_H, \tilde{q}_H, \tilde{c}_{X_t}, \tilde{q}_{X_t}\}$, which can be obtained by equating the corresponding partial derivatives of $\Phi^{(\tau)}$ to zero. In doing so, as $\tau \rightarrow 0$, we get that $\tilde{c}_H = 0$, $\tilde{c}_{X_t} = 0$, $c_H = \mathbb{E}\{|H|^2\}$, $c_{X_t} = \mathbb{E}\{|X_t|^2\}$, and the other parameters $\{q_H, q_{X_t}, \tilde{q}_H, \tilde{q}_{X_t}\}$ are given by (20) in Proposition 1. Let $\text{mse}_H = c_H - q_H$ and $\text{mse}_{X_t} = c_{X_t} - q_{X_t}$. After taking these into account, we obtain the RS expression of the average free entropy as

$$\Phi = \alpha \sum_t \beta_t \left(\sum_b \int Dv \Psi_b(V_t) \log \Psi_b(V_t) \right)$$

$$- \alpha I(H; Z_H | \tilde{q}_H) - \sum_{t=1}^2 \beta_t I(X_t; Z_{X_t} | \tilde{q}_{X_t})$$

$$+ \alpha (c_H - q_H) \tilde{q}_H + \sum_{t=1}^2 \beta_t (c_{X_t} - q_{X_t}) \tilde{q}_{X_t}. \quad (32)$$

where we have defined $V_t \triangleq \sqrt{q_H q_{X_t}} v$, $\Psi_b(V_t)$ is given by (22), and the notation $I(X, Z | q)$ is used to denote the mutual information between X and Z with a Gaussian scalar channel $Z = \sqrt{q}X + W$ and $W \sim \mathcal{N}_C(0, 1)$. Note that the parameters $\{q_H, q_{X_t}, \tilde{q}_H, \tilde{q}_{X_t}\}$ given by (20) are saddle points of (32).

REFERENCES

- [1] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.
- [2] E. G. Larsson, F. Tufvesson, O. Edfors, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.
- [3] J. G. Andrews, S. Buzzi, W. Choi, S. Hanly, A. Lozano, A.C.K. Soong, and J. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, June 2014.
- [4] J. Singh, O. Dabeer, and U. Madhow, "On the limits of communication with low-precision analog-to-digital conversion at the receiver," *IEEE Trans. Commun.*, vol. 57, no. 12, pp. 3629–3639, Dec. 2009.
- [5] J. Mo and R. W. Heath Jr, "Capacity analysis of one-bit quantized MIMO systems with transmitter channel state information," preprint, 2014. [Online]. Available: <http://arxiv.org/abs/1410.7353>
- [6] K. Nakamura and T. Tanaka, "Performance analysis of signal detection using quantized received signals of linear vector channel," in *Proc. Inter. Symp. Inform. Theory and its Applications (ISITA)*, Auckland, New Zealand, Dec. 2008.
- [7] A. Mezghani and J. Nossek, "Belief propagation based MIMO detection operating on quantized channel output," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, Austin, TX, 13–18 June 2010, pp. 2113–2117.

- [8] C. Risi, D. Persson, and E. G. Larsson, "Massive MIMO with 1-bit ADC," preprint, 2014. [Online]. Available: <http://arxiv.org/abs/1404.7736>.
- [9] S. Wang, Y. Li, and J. Wang, "Multiuser detection in massive spatial modulation MIMO with low-resolution ADCs," *IEEE Trans. Wireless Commun.*, 2015.
- [10] J. Choi, D. J. Love, D. R. Brown III, M. Boutin, "Distributed reception with spatial multiplexing: MIMO systems for the internet of things," preprint, 2014. [Online]. Available: <http://arxiv.org/abs/1409.7850>
- [11] Y. Xu, Y. Kabashima, and Lenka Zdeborová, "Bayesian signal reconstruction for 1-bit compressed sensing," *J. Stat. Mech.*, no. 11, p. P11015, 2014.
- [12] K. Takeuchi, R. R. Müller, M. Vehkaperä, and T. Tanaka, "On an achievable rate of large Rayleigh block-fading MIMO channels with no CSI," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6517–6541, Oct. 2013.
- [13] J. Ma and L. Ping, "Data-aided channel estimation in large antenna systems," *IEEE Trans. Signal Processing*, vol. 62, no. 12, pp. 3111–3124, June 2014.
- [14] C.-K. Wen, Y. Wu, K.-K. Wong, R. Schober, and P. Ting, "Performance limits of massive MIMO systems based on Bayes-optimal inference," in *IEEE Int. Conf. Commun. (ICC)*, London, UK, 2015.
- [15] J. T. Parker and P. Schniter and V. Cevher, "Bilinear generalized approximate message passing," *IEEE Trans. Sig. Proc.*, vol. 62, no. 22, pp. 5839–5853, Nov. 2014.
- [16] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1994.
- [17] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1 ed., 1988.
- [18] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing," *Proc. Nat. Acad. Sci.*, vol. 106, no. 45, pp. 18 914–18 919, 2009.
- [19] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," *ArXiv e-prints:1010.5141v2 [cs.IT]*, 2010.
- [20] F. Krzakala, M. Mézard, and L. Zdeborová, "Phase diagram and approximate message passing for blind calibration and dictionary learning," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, Istanbul, Turkey, July 2013, pp. 659–663.
- [21] Y. Kabashima, F. Krzakala, M. Mézard, A. Sakata, and L. Zdeborová, "Phase transitions and sample complexity in Bayes-optimal matrix factorization," preprint 2014. [Online]. Available: <http://arxiv.org/abs/1402.1298>.
- [22] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing: An Introduction*. ser. Number 111 in Int. Series on Monographs on Physics. Oxford U.K.: Oxford Univ. Press, 2001.
- [23] T. Tanaka, "A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors," *IEEE Trans. Inf. Theory*, vol. 48, no. 11, pp. 2888–2910, Nov. 2002.
- [24] A. L. Moustakas, S. H. Simon, and A. M. Sengupta, "MIMO capacity through correlated channels in the presence of correlated interferers and noise: a (not so) large N analysis," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2545–2561, Oct. 2003.
- [25] D. Guo and S. Verdú, "Randomly spread CDMA: asymptotics via statistical physics," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 1982–2010, Jun. 2005.
- [26] R. R. Müller, "Channel capacity and minimum probability of error in large dual antenna array systems with binary modulation," *IEEE Trans. Sig. Proc.*, vol. 51, no. 11, pp. 2821–2828, Nov. 2003.
- [27] C.-K. Wen and K.-K. Wong, "Asymptotic analysis of spatially correlated MIMO multiple-access channels with arbitrary signaling inputs for joint and separate decoding," *IEEE Trans. Inf. Theory*, vol. 53, no. 1, pp. 252–268, Jan. 2007.
- [28] A. Hatabu, K. Takeda, and Y. Kabashima, "Statistical mechanical analysis of the Kronecker channel model for multiple-input multiple-output wireless communication," *Phys. Rev. E*, vol. 80, pp. 061 124(1–12), 2009.
- [29] M. A. Girnyk, M. Vehkaperä, L. K. Rasmussen, "Large-system analysis of correlated MIMO channels with arbitrary signaling in the presence of interference," *IEEE Trans. Wireless Commun.*, vol. 13, no. 4, pp. 1536–1276, Apr. 2014.